

Räumliche Ökonometrie

Hans-Friedrich-Eckey, Reinhold Kosfeld und Matthias Türck, Kassel

Sind die Untersuchungseinheiten in einem statistisch-ökonometrischen Modell Regionen, so spricht man von einer räumlichen Ökonometrie ("Spatial Econometrics"). Diese Modelle lassen die Schätzung von Einflüssen zwischen den Regionen (Spillover-Effekte) zu. Damit sind breite Anwendungsmöglichkeiten sowohl in der Forschung als auch der betrieblichen Praxis (z. B. Marktforschung) gegeben. Die räumliche Ökonometrie hat in den letzten zwanzig Jahren einen erheblichen methodischen Fortschritt erfahren und sich zu einem eigenständigen Gebiet innerhalb der Ökonometrie entwickelt. Einer ihrer Hauptaspekte gilt dabei der Abhängigkeit der Störgrößen von regionalen Untersuchungseinheiten, die wir als räumliche Autokorrelation bezeichnen. Ihr Vorliegen bei einer Kleinst-Quadrate-Schätzung führt zu einer Verzerrtheit der Regressionskoeffizienten oder einer Ungültigkeit von Signifikanztests. Der vorliegende Beitrag erläutert das Konzept der räumlichen Autokorrelation sowie die Diagnose räumlicher Autokorrelation in einem linearen Regressionsmodell. Darüber hinaus wird dargestellt, wie eine räumliche Abhängigkeit in einem Regressionsmodell berücksichtigt werden kann.

Prof. Dr. Hans-Friedrich Eckey ist Leiter des Fachgebiets "Empirische Wirtschaftsforschung" an der Universität Kassel. Seine bevorzugten Forschungsschwerpunkte sind Ökonometrie, regionale Strukturpolitik, Raumordnungspolitik, Verkehrsökonomie und Konjunkturpolitik.

Prof. Dr. Reinhold Kosfeld vertritt das Fachgebiet "Statistik" an der Universität Kassel. Seine Forschungsschwerpunkte erstrecken sich über die Bereiche ökonometrische Methoden, multivariate statistische Verfahren, räumliche Ökonometrie und empirische Wirtschaftsforschung.

Matthias Türck, M. A., ist Wissenschaftlicher Mitarbeiter am Fachgebiet "Empirische Wirtschaftsforschung" an der Universität Kassel. Seine Arbeitsschwerpunkte liegen in den Bereichen Ökonometrie, Statistik und Regionalökonomie.

1. Konzept räumlicher Autokorrelation

Die **räumliche Ökonometrie** ("Spatial Econometrics") umfasst Regressionsmodelle, die bei der Schätzung die Nachbarschaft von Regionen einbeziehen. Bevor wir auf die am häufigsten verwendeten Modelle eingehen, soll das Konzept der räumlichen Abhängigkeit verdeutlicht werden.

Wir gehen von dem gewöhnlichen Regressionsmodell

$$(1) \quad \mathbf{y} = \beta_0 \cdot \mathbf{i} + \beta_1 \cdot \mathbf{x}_1 + \beta_2 \cdot \mathbf{x}_2 + \dots + \beta_p \cdot \mathbf{x}_p + \mathbf{u}$$

aus. Hierin bezeichnen \mathbf{y} einen $n \times 1$ -Vektor der abhängigen Variablen, \mathbf{x}_i einen $n \times 1$ -Vektor der i -ten erklärenden Variablen, \mathbf{u} einen $n \times 1$ -Vektor der Störgröße und \mathbf{i} den $n \times 1$ -Einsvektor. Bei Gültigkeit der Standardannahmen – keine Heteroskedastizität und Autokorrelation liegen vor – können die Regressionskoeffizienten $\beta_0, \beta_1, \dots, \beta_p$ mit der Kleinst-Quadrate-Methode (**OLS-Methode**) geschätzt werden (vgl. Eckey/Kosfeld/Dreger 2004, S. 19 ff.).

Die Fragestellungen für ein Regionsmodell mit regionalen Untersuchungseinheiten können vielfältig sein:

Fallbeispiel 1: Für einen Massenartikel, den viele Anbieter herstellen, soll die Wirkung von Absatzveränderungen auf den Preis berechnet werden. Die Daten beziehen sich auf die Nielsen-Gebiete, einer Einteilung des Bundesgebietes in acht regionale Einheiten, die auf das amerikanische Marktforschungsinstitut AC Nielsen zurückgeht. Dieses Beispiel werden wir im Folgenden fortführen.

Fallbeispiel 2: Eine absolute Konvergenz impliziert, dass ärmere Regionen schneller wachsen als reichere. Diese Fragestellung wird für deutsche Regionen untersucht, indem wir uns die durchschnittliche Wachstumsrate des Bruttoinlandsproduktes über das Bruttoinlandsprodukt des Basisjahres erklären lassen. Ein signifikantes Steigungsmaß würde auf eine absolute Konvergenz hindeuten.

Fallbeispiel 3: Ein international tätiges Unternehmen möchte herausfinden, wie stark die Anzahl der Filialen pro Einwohner den Umsatz pro Einwohner beeinflusst. Die Werte beider Variablen liegen auf Länderebene vor.

Verwendet man regionale Untersuchungseinheiten in einem Regressionsmodell, dann ist häufig eine räumliche Autokorrelation vorhanden. Diese kann beispielsweise durch Spillover-Effekte (Überschwappeneffekte) beim Wissen oder das Area-Unit-Problem hervorgerufen werden. Spillover-Effekte entstehen dadurch, dass sich vor allem "Tacit Knowledge" – Wissen, das insbesondere durch persönliche Kontakte übermittelt wird – durch regionale Netzwerke ausbreitet, die sich regionsübergreifend bilden. Neuere Untersuchungen belegen, dass Beziehungen über Telekommunikationsmedien weit weniger intensiv sind wie persönliche Kontakte (Bretschger 1999, S. 252). Insofern ist die Reichweite regionaler Netzwerke i. d. R. auf einige benachbarte Regionen beschränkt, so dass eine räumliche Abhängigkeit angrenzender Gebiete entsteht.

Das Area-Unit-Root-Problem ist relevant, wenn mit Verwaltungseinheiten und nicht funktionalen Gebieten gearbeitet wird. Während Verwaltungseinheiten, z. B. Kreise und Bundesländer, meistens nicht die räumlichen Verflechtungen berücksichtigen, werden funktionale Räume z. B. aufgrund von Pendlerströmen (= regionale Arbeitsmärkte) abgegrenzt. Funktionsgebiete spiegeln damit die wirtschaftlichen Aktivitäten im Raum wider und stellen relativ eigenständige wirtschaftliche Teilräume dar (Keilbach 2000, S. 120 f.). Damit "zerschneidet" eine Verwaltungseinheit funktionale Gebiete, so dass daraus eine Abhängigkeit zwischen angrenzenden Verwaltungsregionen resultiert.

Bei Vorliegen räumlicher Autokorrelation sind die Residuen angrenzender Regionen nicht mehr unabhängig voneinander. Die Kovarianz der Regionen i und j weicht dann zumindest teilweise von null ab:

$$(2) \quad \text{Cov}(u_i, u_j) \neq 0 \quad \text{für } i \neq j.$$

Damit ist die OLS-Schätzung nicht erwartungstreu oder die Signifikanztests verlieren ihre Gültigkeit. Bei einer fehlenden Erwartungstreue stimmt der durchschnittliche Schätzer nicht mit seinem Erwartungswert überein. Er ist entweder nach oben oder unten verzerrt. Eine verzerrt geschätzte Varianz der Regressionskoeffizienten führt zu einer Ungültigkeit der Signifikanztests. Die einzelnen Einflüsse lassen sich nicht mehr testen. Unabhängig von der Auswirkung räumlicher Autokorrelation muss die regionale Abhängigkeit im Modell berücksichtigt werden.

2. Moran-Koeffizient

Wie kann das Vorliegen räumlicher Autokorrelation diagnostiziert werden? Für das Konzept der regionalen Autokorrelation ist die **Gewichtungsmatrix \mathbf{W}^*** maßgeblich. Diese basiert entweder auf den Distanzen zwischen den Regionen oder einer binären 0/1-Matrix. Wir gehen vom letzteren Fall aus, für den die Gewichtungsmatrix wesentlich leichter erstellt werden kann. Bei dieser binären Kodierung gibt die Gewichtungsmatrix an, ob zwei Regionen aneinander angrenzen:

$$(3) \quad w_{ij}^* = \begin{cases} 1, & \text{falls } i \text{ und } j \text{ eine gemeinsame Grenze haben und } i \neq j \text{ ist} \\ 0, & \text{sonst} \end{cases}$$

Diese Gewichtungsmatrix \mathbf{W}^* wird in der Regel so normiert, dass die Summe der Zeilenelemente eins ergibt. Sie erhält man, indem jedes Element von \mathbf{W}^* durch die jeweilige Zeilensumme dividiert wird:

$$(4) \quad w_{ij} = \frac{w_{ij}^*}{\sum_{j=1}^n w_{ij}^*}.$$

Häufig liegt eine positive räumliche Autokorrelation vor. Dann weisen die Residuen in den Nachbarregionen meistens das gleiche Vorzeichen auf. Das Produkt der Residuen in Region j mit dem arithmetischen Mittel in den Nachbarregionen (**Spatial-Lag** in den Residuen),

$$(5) \quad \mathbf{W} \cdot \hat{\mathbf{u}} = \sum_{i=1}^n w_{ji} \cdot \hat{u}_i,$$

ist dann häufig positiv:

$$(6) \quad \hat{u}_j \cdot \sum_{i=1}^n w_{ji} \cdot \hat{u}_i.$$

Der **Moran-Koeffizient** (Moran's I) bezieht die Summe dieser Kreuzprodukte auf die Quadratsumme der Residuen (vgl. Cliff/Ord 1973, S. 87 ff. und Upton/Fingleton 1985, S. 337 f.):

$$(7) \quad I = \frac{\sum_{j=1}^n \hat{u}_j \cdot \left(\sum_{i=1}^n w_{ji} \cdot \hat{u}_i \right)}{\sum_{j=1}^n \hat{u}_j^2} = \frac{\hat{\mathbf{u}}' \mathbf{W} \cdot \hat{\mathbf{u}}}{\hat{\mathbf{u}}' \cdot \hat{\mathbf{u}}}$$

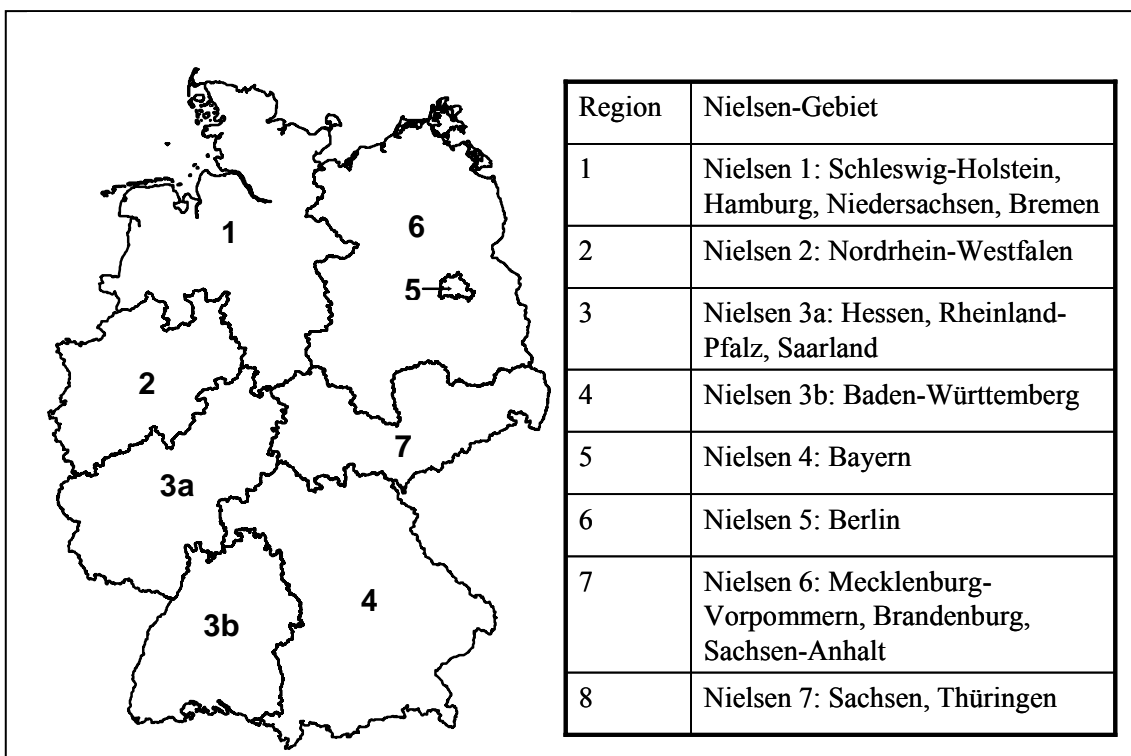
Da die Quadratsumme der Residuen stets positive Werte annimmt, bestimmt der Zähler das Vorzeichen des Moran-Koeffizienten. Ist die Zählergröße und damit der gesamte Bruch größer als null, dann weist Moran's I auf eine positive Autokorrelation hin. Zu beachten ist, dass der Erwartungswert vom Moran-Koeffizienten von \mathbf{X} – der Matrix von den unabhängigen Variablen – abhängt und beispielsweise, wenn nur die Scheinvariable als einziger Regressor vorhanden ist, bei $-1/n$ liegt. Insofern weist ein Moran-Koeffizient, der in der Nähe von null oder leicht negativ ist, auf keine Autokorrelation hin. Eine negative regionale Autokorrelation ist nur für wenige Fragestellungen denkbar, beispielsweise wenn eine Region durch Umweltschäden vom Nachbargebiet geschädigt wird.

Wie gezeigt werden kann (vgl. Cliff/Ord 1973, S. 99 f. und Cliff/Ord 1981, S. 66 ff.), folgt Moran's I nach einer Standardisierung:

$$(8) \quad Z = \frac{I - E(I)}{\sqrt{\text{Var}(I)}}$$

unter der Nullhypothese unabhängiger, identisch verteilter Störgrößen approximativ einer Standardnormalverteilung. Somit kann Moran's I auf räumliche Autokorrelation getestet werden (s. Anselin/Florax 1995, S. 25). Auf die genaue Berechnungsformel soll aufgrund der komplexen Ausdrücke nicht eingegangen werden. Der interessierte Leser findet sie in Upton/Fingleton 1985, S. 337 ff. Liegt eine substantielle Autokorrelation vor, dann muss die räumliche Abhängigkeit durch eines der in den Abschnitten 3 und 4 erläuterten Modelle "eingefangen" werden.

Abb. 1: Nielsen-Gebiete



Verdeutlichen wir uns die Berechnung des Moran-Koeffizienten an einem Beispiel. Wir gehen von den Nielsen-Gebieten aus, deren räumliche Anordnung in Abb. 1 eingezeichnet ist. Die Regionen 1 und 2 grenzen aneinander an, weshalb wir in die Gewichtungsmatrix \mathbf{W}^* eine 1 für die Elemente (2,1) und (1,2) eintragen. Entsprechend gehen wir mit den anderen Regionen vor. Auf der Hauptdiagonale stehen jeweils Nullen:

$$\mathbf{W}^* = \begin{matrix} & \begin{matrix} 1 & 2 & 3a & 3b & 4 & 5 & 6 & 7 \end{matrix} \\ \begin{matrix} 0 \\ 1 \\ 1 \\ 0 \\ 0 \\ 0 \\ 1 \\ 1 \end{matrix} & \begin{bmatrix} 1 & 0 & 1 & 0 & 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 1 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 1 & 0 & 1 & 0 & 0 \end{bmatrix} \end{matrix} \begin{matrix} 1 \\ 2 \\ 3a \\ 3b \\ 4 \\ 5 \\ 6 \\ 7 \end{matrix}.$$

Diese Gewichtungsmatrix \mathbf{W}^* wird normiert, indem man die Elemente durch die Zeilensumme dividiert. Die Summe von Zeile 1 beläuft sich beispielsweise auf

$$\sum_{j=1}^8 w_{1j}^* = 0 + 1 + 1 + 0 + 0 + 0 + 1 + 1 = 4,$$

so dass die Elemente in der ersten Zeile der normierten Matrix entweder 0 oder 1/4 betragen. Entsprechend hat die komplette normierte Gewichtungsmatrix folgendes Aussehen:

$$\mathbf{W} = \begin{matrix} & \begin{matrix} 1 & 2 & 3a & 3b & 4 & 5 & 6 & 7 \end{matrix} \\ \begin{matrix} 0 \\ 1/2 \\ 1/5 \\ 0 \\ 0 \\ 0 \\ 1/3 \\ 1/4 \end{matrix} & \begin{bmatrix} 1/4 & 0 & 1/4 & 0 & 0 & 0 & 1/4 & 1/4 \\ 0 & 1/2 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1/5 & 0 & 1/5 & 1/5 & 0 & 0 & 1/5 \\ 0 & 0 & 1/2 & 0 & 1/2 & 0 & 0 & 0 \\ 0 & 0 & 1/3 & 1/3 & 0 & 0 & 0 & 1/3 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1/1 & 0 \\ 1/3 & 0 & 0 & 0 & 0 & 1/3 & 0 & 1/3 \\ 1/4 & 0 & 1/4 & 0 & 1/4 & 0 & 1/4 & 0 \end{bmatrix} \end{matrix} \begin{matrix} 1 \\ 2 \\ 3a \\ 3b \\ 4 \\ 5 \\ 6 \\ 7 \end{matrix}.$$

Wir gehen von den in Tab. 1 dargestellten Daten einer Preis-Absatz-Funktion für die Nielsen-Gebiete eines Massenartikels, den viele Produzenten herstellen, aus. Der Preis ist dabei in Euro angegeben. Der Absatz wird als Durchschnittswert pro Einwohner und Jahr gemessen.

Tab. 1: Beispielhafte Daten

Region (i)	1	2	3	4	5	6	7	8
Preis (Y)	12	15	12	15	15	14	10	9
Absatz (X)	8	7	14	6	6	3	10	16

Die OLS-Schätzung der Gleichung (1) liefert ein Modell, dessen Güte positiv zu bewerten ist. Der p-Wert des F-Tests, der in der letzten Zeile von Tab. 2 angegeben ist,

unterschreitet nämlich das Signifikanzniveau von 5 %. Der Anteil der erklärten Varianz liegt bei 62,7 %. Die p-Werte der Regressionskoeffizienten sind ebenfalls beide geringer als 0,05, so dass beide Regressionskoeffizienten signifikant sind, also nicht nur zufällig von null abweichen. Das absolute Glied liegt bei 16,519 und ist schlecht zu interpretieren. Es würde einen Preis darstellen, der bei einem Absatz von null Einheiten erzielt worden wäre. Das Steigungsmaß beträgt -0,431. Erhöht sich der Absatz um ein Stück je Einwohner, dann sinkt also der Preis um durchschnittlich 0,431 Euro.

Tab. 2: OLS-Schätzung ohne Spatial-Lags

Variable	Regressionskoeffizient	Standardfehler	t-Wert	p-Wert
Absolutes Glied	16,519	1,310	12,613	0,000
Absatz	-0,431	0,136	-3,176	0,019
$R^2 = 0,627$; $F = 10,085$; $p = 0,019$				

Mit Hilfe der Regressionskoeffizienten lassen sich der geschätzte y-Wert für die erste Region:

$$\hat{y}_1 = 16,519 - 0,431 \cdot x_1 = 16,519 - 0,431 \cdot 8 = 13,071$$

und damit das OLS-Residuum:

$$\hat{u}_1 = y_1 - \hat{y}_1 = 12 - 13,071 = -1,071$$

bestimmen. Entsprechend wird mit den übrigen Regionen verfahren, so dass sich folgender Residualvektor ergibt:

$$\hat{\mathbf{u}}' = [-1,071 \quad 1,498 \quad 1,515 \quad 1,067 \quad 1,067 \quad -1,226 \quad -2,209 \quad -0,623].$$

Dann beträgt das Matrixprodukt zwischen \mathbf{W} und $\hat{\mathbf{u}}$ (Spatial-Lag in den Residuen):

$$(\mathbf{W}\hat{\mathbf{u}})' = [0,442 \quad 0,219 \quad 0,385 \quad 1,288 \quad 0,650 \quad -2,212 \quad -0,976 \quad -0,177].$$

Die Nachbarregionen von Region 1 haben damit beispielsweise ein durchschnittliches Residuum in Höhe von 0,442. Als endgültigen Wert für den Moran-Koeffizienten erhält man schließlich:

$$I = \frac{\hat{\mathbf{u}}' \cdot \mathbf{W} \cdot \hat{\mathbf{u}}}{\hat{\mathbf{u}}' \cdot \hat{\mathbf{u}}} = \frac{7,912}{14,734} = 0,537.$$

Der Moran-Koeffizient, der i. d. R. zwischen minus und plus eins liegt, weist hier eine deutlich positive Autokorrelation aus. Mit Hilfe seines Erwartungswertes und seiner Varianz lässt sich der z-Wert ermitteln [vgl. Formel (8)]. Da es sich um eine konkrete Ausprägung und keine Zufallsvariable handelt, verwenden wir einen Kleinbuchstaben:

$$z = \frac{0,537 - (-0,141)}{\sqrt{0,0631}} = 2,699.$$

Das entsprechende 95 %-Quantil (= kritischer Wert) bei einem zweiseitigen Test kann den Tabellen der Standardnormalverteilung entnommen werden, wie sie in Statistiklehrbüchern abgedruckt sind. Der kritische Wert liegt bei 1,96 und ist kleiner als 2,699, so dass die Nullhypothese keiner Autokorrelation abgelehnt wird. Das vorhandene Modell ist damit nicht adäquat. Wir müssen einen Spatial-Lag in der exogenen Variablen, also dem Absatz, der abhängigen Variablen (Spatial-Lag-Modell) oder dem Fehlerterm (Spatial-Error-Modell) berücksichtigen.

3. Spatial-Lag in den exogenen Variablen

Bei Vorliegen räumlicher Autokorrelation lassen sich Spatial-Lags in einer oder mehreren x-Variablen in Gleichung (1) einbeziehen. Dieser Ansatz unterstellt, dass exogene Variablen nicht nur die abhängige Variable Y in der gleichen Region, sondern auch in den Nachbarregionen beeinflussen. Das entsprechende Modell, bei dem Spatial-Lags in allen exogenen Variablen einfließen, lautet (Spatial-Cross-Regressive-Modell):

$$(9) \quad \mathbf{y} = \beta_0 \cdot \mathbf{i} + \beta_1 \cdot \mathbf{x}_1 + \dots + \beta_p \cdot \mathbf{x}_p + \beta_{p+1} \cdot \mathbf{W} \cdot \mathbf{x}_1 + \dots + \beta_{p+p} \cdot \mathbf{W} \cdot \mathbf{x}_p + \mathbf{u}.$$

Spatial-Lags in den exogenen Variablen stellen die einfachste Form der Berücksichtigung einer räumlichen Autokorrelation dar. Das Modell kann nämlich weiterhin mit der OLS-Methode geschätzt werden, soweit die Störgrößen insbesondere keine räumliche Autokorrelation aufweisen. Als Diagnosemöglichkeit bietet sich weiterhin der Moran-Koeffizient für die OLS-Residuen an [vgl. Formel (7) und (8)].

Wir wollen für das vorliegende Beispiel einen Spatial-Lag beim Absatz einbeziehen:

$$\text{Preis}_i = \beta_0 + \beta_1 \cdot \text{Absatz}_i + \beta_2 \cdot \mathbf{W} \cdot \text{Absatz}_i + u_i.$$

Dieser wird als Matrixprodukt zwischen der Gewichtungsmatrix \mathbf{W} und den Absätzen, die im Vektor

$$\mathbf{x}'_2 = [8 \quad 7 \quad 14 \quad 6 \quad 6 \quad 3 \quad 10 \quad 16]$$

angeordnet sind, ermittelt:

$$(\mathbf{W}\mathbf{x}_2)' = [11,75 \quad 11 \quad 8,6 \quad 10 \quad 12 \quad 10 \quad 9 \quad 9,5].$$

Der Absatz in den Nachbarregionen von Region 1 liegt also im Mittel bei 11,75 Stück pro Einwohner. Damit führen wir eine OLS-Schätzung unter Verwendung des Vektors

$$\mathbf{y}' = [12 \quad 15 \quad 12 \quad 15 \quad 15 \quad 14 \quad 10 \quad 9]$$

und der Matrix

$$\mathbf{X}' = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 8 & 7 & 14 & 6 & 6 & 3 & 10 & 16 \\ 11,75 & 11 & 8,6 & 10 & 12 & 10 & 9 & 9,5 \end{bmatrix}$$

durch. Unter Verwendung folgender Zwischenergebnisse,

$$(\mathbf{X}'\mathbf{X})^{-1} = \begin{bmatrix} 18,3664 & -0,3028 & -1,5239 \\ -0,3028 & 0,0107 & 0,0205 \\ -1,5239 & 0,0205 & 0,1314 \end{bmatrix}, \quad \mathbf{X}'\mathbf{y} = \begin{bmatrix} 102 \\ 835 \\ 1.054,7 \end{bmatrix},$$

ergibt sich als Vektor der Regressionskoeffizienten:

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{y} = \begin{bmatrix} \hat{\beta}_1 = 13,228 \\ \hat{\beta}_2 = -0,387 \\ \hat{\beta}_3 = 0,284 \end{bmatrix}.$$

Wie sind die beiden Regressionskoeffizienten für den Absatz der eigenen Region und des angrenzenden Gebiets zu interpretieren? Erhöht sich der Absatz in der eigenen Region um ein Stück je Einwohner bei Konstanzhaltung dieser Variablen in den angrenzenden Regionen, dann sinkt der Preis in der eigenen Region um durchschnittlich 0,387 Euro. In den Nachbarregionen führt eine durchschnittliche Absatzzunahme um ein Stück je Einwohner – der Absatz in der eigenen Region verändert sich nicht – dagegen zu einer mittleren Preiserhöhung um 0,284 Euro in der eigenen Region. Der

Verkauf in den angrenzenden Gebieten wird relativ attraktiver, weshalb der Preis in der eigenen Region steigt.

Tab. 3: OLS-Schätzung mit einem Spatial-Lag beim Absatz

Variable	Regressionskoeffizient	Standardfehler	t-Wert	p-Wert
Absolutes Glied	13,228	7,202	1,837	0,126
Absatz	-0,387	0,174	-2,226	0,077
Spatial-Lag beim Absatz	0,284	0,609	0,466	0,661
$R^2 = 0,642$; $F = 4,493$; $p = 0,076$				

Mit den Regressionskoeffizienten kann der F-Test berechnet werden, der allerdings hier nur ein schwach signifikantes Ergebnis (Signifikanzniveau von 10 %) liefert (vgl. Tab. 3). Auch die Regressionskoeffizienten sind bei einem Signifikanzniveau von 5 % alle nicht signifikant, ihr p-Wert überschreitet also die 0,05. Insofern sind die Regressionskoeffizienten nicht statistisch abgesichert.

Hat die Einbeziehung des Spatial-Lags beim Absatz die räumliche Autokorrelation beseitigt? Wir berechnen wiederum den Moran-Koeffizienten, der allerdings mit einem Wert von

$$I = \frac{\hat{\mathbf{u}}' \cdot \mathbf{W} \cdot \hat{\mathbf{u}}}{\hat{\mathbf{u}}' \cdot \hat{\mathbf{u}}} = \frac{6,344}{14,121} = 0,449$$

immer noch eine erhebliche positive Autokorrelation anzeigt, die weiterhin signifikant ist. Die Schätzung ist damit ungeeignet. Das Modell eines Spatial-Lags in den exogenen Variablen führt uns also nicht weiter, so dass ein Spatial-Lag- oder Spatial-Error-Modell anzuwenden ist.

4. Spatial-Lag- und Spatial-Error-Modell

Eine räumliche Autokorrelation in einem mit der OLS-Methode geschätzten Modell lässt sich auch über einen Spatial-Lag in der abhängigen Variablen (Spatial-Lag-Modell):

$$(10) \quad \mathbf{y} = \rho \cdot \mathbf{W} \cdot \mathbf{y} + \beta_0 \cdot \mathbf{i} + \beta_1 \cdot \mathbf{x}_1 + \beta_2 \cdot \mathbf{x}_2 + \dots + \beta_p \cdot \mathbf{x}_p + \mathbf{u}$$

oder im Fehlerterm (Spatial-Error-Modell):

$$(11) \quad \mathbf{y} = \beta_0 \cdot \mathbf{i} + \beta_1 \cdot \mathbf{x}_1 + \beta_2 \cdot \mathbf{x}_2 + \dots + \beta_p \cdot \mathbf{x}_p + \lambda \cdot \mathbf{W} \cdot \mathbf{u} + \boldsymbol{\varepsilon}$$

modellieren.

Ein Spatial-Lag in der abhängigen Variablen ist genauso zu interpretieren wie der Spatial-Lag in den exogenen Variablen. Er stellt einen Durchschnittswert von Y in den angrenzenden Regionen dar. Der Spatial-Lag im Fehlerterm [vgl. Formel (11)] lässt sich durch zufällige Schocks, die Nichtberücksichtigung relevanter Einflussgrößen, insbesondere in Form von Spatial-Lags exogener Variabler, oder das eingangs erläuterte Area-Unit-Problem begründen. Eine Entscheidungshilfe bei der Wahl eines Spatial-Error- oder Spatial-Lag-Modells geben die Lagrange-Multiplier-Tests (Florax/Rey 1995, S. 119 ff.). Beide Modelle lassen zusätzlich die Berücksichtigung von Spatial-Lags in den exogenen Variablen zu (vgl. Lauridsen 2002 u. Kosfeld/Lauridsen 2004):

Da der Spatial-Lag in Y oder U nicht mehr exogen vorgegeben ist wie in den Modellen aus Abschnitt 3, darf eine OLS-Schätzung aufgrund von Verzerrungen der Regressionskoeffizienten oder einer Invalidierung der Signifikanztests (s. Anselin 1988, S. 57 ff.) nicht durchgeführt werden. Hier findet dann beispielsweise die **Maximum-Likelihood-Schätzung** Anwendung, die allerdings zusätzlich normalverteilte Störgrößen voraussetzt. Die relativ komplexen Formeln seien an dieser Stelle nicht hergeleitet. Der interessierte Leser findet sie in Upton/Fingleton (1985) oder Anselin (1988). Wir veranschaulichen die Schätzung dagegen an einem Beispiel.

Da sich der Spatial-Lag beim Absatz im OLS-Modell (vgl. Abschnitt 3) nicht als signifikant erwiesen hat, gehen wir von dem ursprünglichen Modell aus. Der Lagrange-Multiplier-Test für das Spatial-Error-Modell ist schwach signifikant (bei einem Signifikanzniveau von 10 %), während der entsprechende Test für das Spatial-Lag-Modell keine Signifikanz ausweist. Insofern würde man das Spatial-Error-Modell berechnen.

Wir wollen zur Demonstration jedoch sowohl das Spatial-Lag- als auch das Spatial-Error-Modell erläutern. Aus didaktischen Gründen beginnen wir mit dem erstgenannten Ansatz, der durch die Gleichung

$$\text{Preis}_i = \rho \cdot \mathbf{W} \cdot \text{Preis}_i + \beta_0 + \beta_1 \cdot \text{Absatz}_i + u_i$$

gegeben ist. Zur Schätzung der Koeffizienten verwenden wir die Maximum-Likelihood-Methode. Hierbei wird ausgehend von der Normalverteilung für die Störgrößen eine Likelihood-Funktion abgeleitet. Der Parameter ρ ist so zu bestimmen, dass er bei gegebenen Werten der abhängigen und unabhängigen Variablen die größte Wahrscheinlichkeit besitzt. Wir gehen dabei in mehreren Schritten vor:

Zuerst berechnet man die Residuen zweier Regressionen, zu deren Unterscheidung wir die auf Anselin (1988, S. 180 ff.) zurückgehenden Bezeichnungen O (für ordinary oder "normal") und L (Modell für die Likelihood-Schätzung) einführen:

$$\text{Preis}_i = \beta_{O0} + \beta_{O1} \cdot \text{Absatz}_i + u_{O_i}$$

und

$$\mathbf{W} \cdot \text{Preis}_i = \beta_{L0} + \beta_{L1} \cdot \text{Absatz}_i + u_{L_i}$$

mit:

$$(\mathbf{W} \cdot \text{Preis})' = [11,5 \quad 12 \quad 13,2 \quad 13,5 \quad 12 \quad 10 \quad 11,667 \quad 12,25].$$

Mit Hilfe der Regressionskoeffizienten dieser Modelle:

$$\hat{\boldsymbol{\beta}}_O = \begin{bmatrix} \hat{\beta}_{O0} = 16,519 \\ \hat{\beta}_{O1} = -0,431 \end{bmatrix} \text{ und } \hat{\boldsymbol{\beta}}_L = \begin{bmatrix} \hat{\beta}_{L0} = 11,002 \\ \hat{\beta}_{L1} = 0,116 \end{bmatrix}$$

werden die Residualvektoren ermittelt ($\hat{\mathbf{u}}_O$ wurde bereits in Abschnitt 2 bestimmt):

$$\hat{\mathbf{u}}'_O = [-1,071 \quad 1,498 \quad 1,515 \quad 1,067 \quad 1,067 \quad -1,226 \quad -2,209 \quad -0,623]$$

und

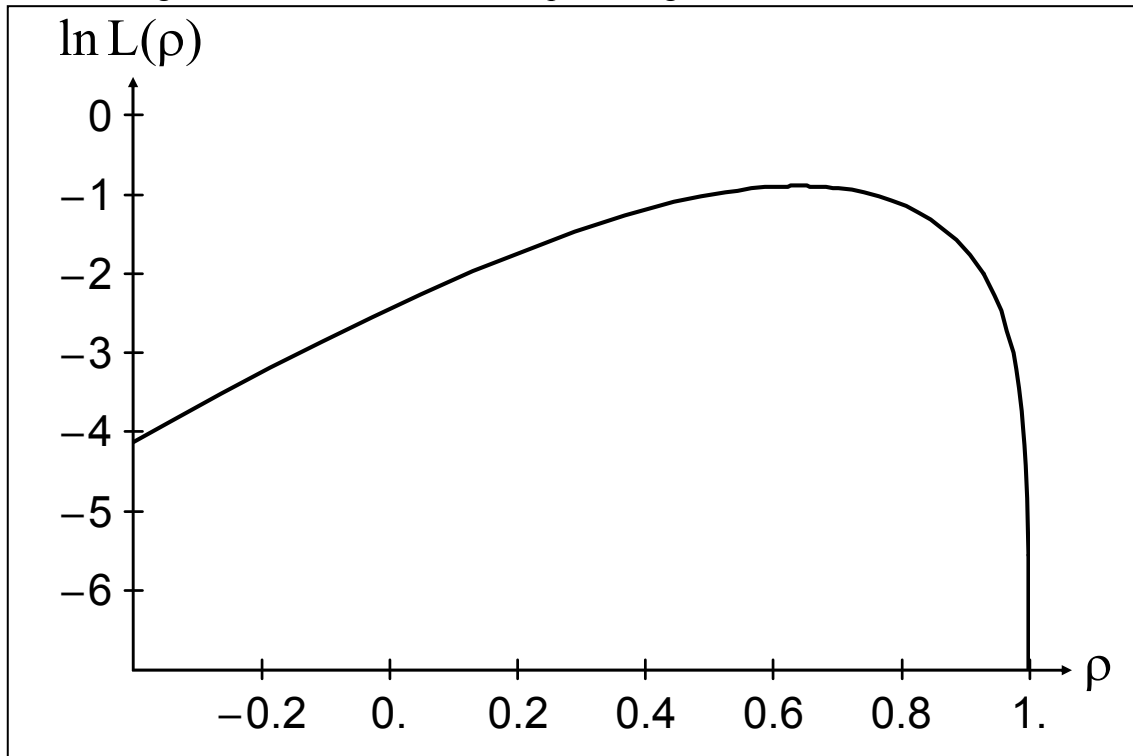
$$\hat{\mathbf{u}}'_L = [-0,428 \quad 0,188 \quad 0,578 \quad 1,804 \quad 0,304 \quad -1,349 \quad -0,493 \quad -0,603].$$

Diese beiden Vektoren benötigt man zur Berechnung des maximalen Wertes der Likelihood-Funktion. Gewöhnlich logarithmiert man die Likelihood-Funktion, da sich das Maximum dann leichter bestimmen lässt. Durch eine Logarithmierung ändert sich nämlich nicht die Position der Extremwerte. Die logarithmierte Likelihood-Funktion, **Log-Likelihood-Funktion**, hat folgendes Aussehen:

$$\ln L(\rho) = -(n/2) \cdot \ln \left[(1/n) \cdot (\hat{\mathbf{u}}_O - \rho \cdot \hat{\mathbf{u}}_L)' \cdot (\hat{\mathbf{u}}_O - \rho \cdot \hat{\mathbf{u}}_L) \right] + \ln |\mathbf{I} - \rho \cdot \mathbf{W}| ,$$

wobei \mathbf{I} die Einheitsmatrix darstellt. Die verschiedenen Funktionswerte der Log-Likelihood-Funktion bei alternativen Parametern ρ sind in Abb. 2 eingezeichnet. Gesucht ist der maximale Funktionswert, der bei 0,639 liegt.

Abb. 2: Log-Likelihood-Funktion beim Spatial-Lag-Modell



Die übrigen beiden Regressionskoeffizienten werden über die Beziehung:

$$\hat{\boldsymbol{\beta}} = \hat{\boldsymbol{\beta}}_O - \rho \cdot \hat{\boldsymbol{\beta}}_L = \begin{bmatrix} 16,519 \\ -0,431 \end{bmatrix} - 0,639 \cdot \begin{bmatrix} 11,002 \\ 0,116 \end{bmatrix} = \begin{bmatrix} \hat{\beta}_0 = 9,489 \\ \hat{\beta}_1 - 0,505 \end{bmatrix}$$

bestimmt. Man erhält dann folgende Regressionsgleichung:

$$\text{Preis}_i = 0,639 \cdot \mathbf{W} \cdot \text{Preis}_i + 9,489 - 0,505 \cdot \text{Absatz}_i + \hat{u}_i .$$

Alle Koeffizienten sind bei einem Signifikanzniveau von 5 % statistisch signifikant. Erhöht sich der durchschnittliche Preis in den Nachbarregionen um einen Euro, dann steigt unter Konstanthaltung des Absatzes der Preis in der eigenen Region im Mittel um 0,639 Euro. Eine Ausweitung des Absatzes um eine Einheit bewirkt dagegen eine Preissenkung um durchschnittlich 0,505 Euro in der gleichen Region, wenn sich der Preis in den Nachbarregionen nicht verändert.

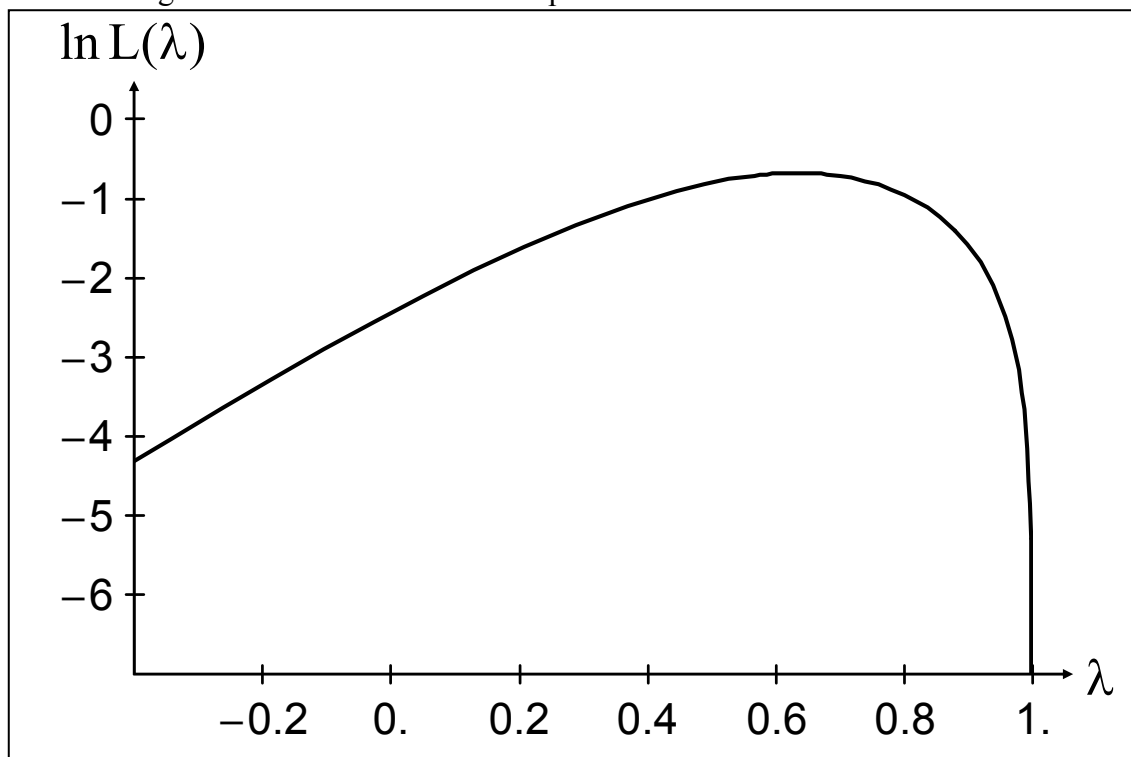
Dieses Spatial-Lag-Modell wäre, wie eingangs erwähnt, aufgrund der Lagrange-Multiplier-Tests nicht auszuwählen. Stattdessen würde man das Spatial-Error-Modell

$$\text{Preis}_i = \beta_0 + \beta_1 \cdot \text{Absatz}_i + \lambda \cdot \mathbf{W} \cdot u_i + \varepsilon_i$$

anwenden. Seine Schätzung ist komplexer als diejenige des Spatial-Lag-Modells, beruht aber auf den gleichen Prinzipien. Auch hier würde man eine Log-Likelihood-Funktion verwenden (vgl. Anselin 1988, S. 182 f.), die sich folgendermaßen darstellen lässt:

$$\ln L(\lambda) = -(n/2) \cdot \ln \left[(1/n) \cdot \hat{\mathbf{u}}_O' \cdot (\mathbf{I} - \lambda \cdot \mathbf{W})' \cdot (\mathbf{I} - \lambda \cdot \mathbf{W}) \cdot \hat{\mathbf{u}}_O \right] + \ln |\mathbf{I} - \lambda \cdot \mathbf{W}| \quad .$$

Abb. 3: Log-Likelihood-Funktion beim Spatial-Error-Modell



Der maximale Funktionswert der Log-Likelihood-Funktion ist einem λ von 0,634 zugeordnet (vgl. Abb. 3). Ausgehend hiervon wird eine GLS-Schätzung (verallgemeinerte Methode der Kleinsten Quadrate) zur Ermittlung der beiden Regressionskoeffizienten durchgeführt:

$$\hat{\boldsymbol{\beta}}_{\text{GLS}} = (\mathbf{X}'\mathbf{B}'\mathbf{B}\mathbf{X})^{-1} \mathbf{X}'\mathbf{B}'\mathbf{B}\mathbf{y} = \begin{bmatrix} \hat{\beta}_0 = 16,859 \\ \hat{\beta}_1 = -0,458 \end{bmatrix}$$

mit

$$\mathbf{B} = \mathbf{I} - \lambda \cdot \mathbf{W} \quad .$$

Falls ein Optimierungskriterium erreicht ist, sind diese Regressionskoeffizienten die gesuchten Werte. Andernfalls wird ein neuer Wert für λ unter Verwendung der GLS-Residuen

$$\hat{\mathbf{u}}_{\text{GLS}} = \mathbf{y} - \mathbf{X} \cdot \hat{\boldsymbol{\beta}}_{\text{GLS}}$$

ermittelt:

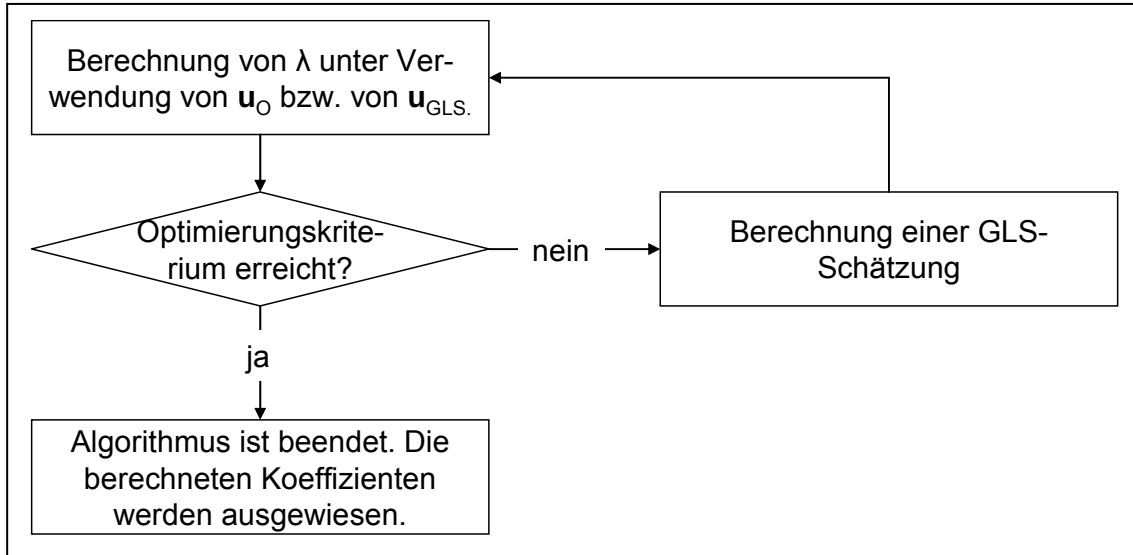
$$\ln L(\lambda) = -(n/2) \cdot \ln \left[(1/n) \cdot \hat{\mathbf{u}}_{\text{GLS}}' \cdot (\mathbf{I} - \lambda \cdot \mathbf{W})' \cdot (\mathbf{I} - \lambda \cdot \mathbf{W}) \cdot \hat{\mathbf{u}}_{\text{GLS}} \right] + \ln |\mathbf{I} - \lambda \cdot \mathbf{W}| \quad .$$

Der Prozess wird so lange fortgeführt, bis das Optimierungskriterium erreicht ist (vgl. Abb. 4). Auf Einzelheiten dieser komplizierten Prozedur soll nicht eingegangen werden. Wir erhalten nach zehn Iterationen folgende Regressionsgleichung:

$$\text{Preis}_i = 16,868 - 0,459 \cdot \text{Absatz}_i + 0,649 \cdot \mathbf{W} \cdot \mathbf{u}_i + \hat{\varepsilon}_i \quad .$$

Alle Koeffizienten weichen signifikant von null ab. Erhöht sich der Absatz um eine Einheit, dann sinkt der Preis in der gleichen Region im Mittel um 0,459 Euro. Der Koeffizient λ gibt zufällige Effekte aus den Nachbarregionen wieder. Er wird nicht interpretiert.

Abb. 4: Iterative Schätzung beim Spatial-Error-Modell



5. Resümee

Die räumlich-ökonometrischen Verfahren finden breite Anwendung. Sie sind relativ anschaulich zu interpretieren, auch wenn die Berechnungsformeln für die in Abschnitt 4 angegebenen Modelle komplex sind. Inzwischen gibt es Spezialsoftware zur Ermittlung der Koeffizienten, beispielsweise das von Anselin entwickelte Programmpaket *SpaceStat* (vgl. <http://www.terraseer.com/products/spacestat.html>, 09/05/05) sowie die von Le Sage zusammengestellte Toolbox für *Matlab* (vgl. <http://www.spatial-econometrics.com>, 09/05/05).

Ein Problem der besprochenen Modelle mit räumlicher Autokorrelation besteht darin, dass die Regressionskoeffizienten für alle Regionen konstant sind, also **räumliche Stationarität** vorausgesetzt wird. Ein weiterer relativ neuer von Fotheringham, Brunsdon und Charlton (2000) entwickelter Ansatz erlaubt die Berechnung von lokal unterschiedlichen Regressionskoeffizienten. Auch hier lassen sich räumliche Effekte modellieren. Simulationsstudien, ob die hier aufgezeigten Modelle der räumlichen Autokorrelation oder die räumlich gewichtete Regression überlegen sind, liegen bis jetzt nicht vor.

Literatur

- Anselin, L., *Spatial Econometrics. Methods and Models*, Dordrecht, Boston, London 1988.
- Anselin, L., R. J. Florax, *Small Sample Properties of Tests for Spatial Dependence in Regression Models*, in: L. Anselin, R. J. Florax (Hrsg.), *New Directions in Spatial Econometrics*, Berlin u. a. 1995, S. 21-52.
- Bretschger, L., *Knowledge Diffusion and the Development of Regions*, *Annals of Regional Science*, Vol. 33(1999), S. 251-268.

- Cliff, A. D., J. K. Ord*, Spatial Autocorrelation, London 1973.
- Cliff, A. D., J. K. Ord*, Spatial Processes. Models and Applications, London 1981.
- Eckey, H.-F., R. Kosfeld, C. Dreger*, Ökonometrie. Grundlagen – Methoden – Beispiele, 3., überarb. und erw. Aufl., Wiesbaden 2004.
- Florax, R. J., S. Rey*, The Impacts of Misspecified Spatial Interaction in Linear Regression Models, in: L. Anselin, R. J. Florax (Hrsg.), New Directions in Spatial Econometrics, Berlin u. a. 1995, S. 21-74.
- Fotheringham, A. S., C. Brunsdon, M. E. Charlton*, Geographically Weighted Regression. The Analysis of Spatial Varying Relationships, Chichester 2000.
- Keilbach, M.*, Spatial Knowledge Spillovers and the Dynamics of Agglomeration and Regional Growth, Diss., Heidelberg, New York 2000.
- Kosfeld, R., J. Lauridsen*, Dynamic Spatial Modelling of Regional Convergence Processes. Empirical Economics, Vol. 29(2004), S. 705-722.
- Lauridsen, J.*, Spatial Autoregressively Distributed Lag Models. Equivalent Forms, Estimation and an Illustrative Commuting Model, Discussion Paper, Department for Statistics and Demography, University of Southern Denmark, Odense 2002.
- Ord, K.*, Estimation Methods for Models of Spatial Interaction. Journal of American Statistical Association, Vol. 70(1975), S. 120-126.
- Upton, G. J., B. Fingleton*, Spatial Data Analysis by Example, Volume I, Chichester u. a. 1985.